



SETTING DATA FREE TO SAVE LIVES

Precision medicine's breakthroughs are achieved with data—so why is hoarding still such a problem, and what can be done to solve it?

When it comes to genomic data, more is always better. The more genomes that are sequenced and shared, the more we learn about the genetic components of disease, and the more likely we are to develop lifesaving treatments.

**“SHARING DATA
IS ESSENTIAL
FOR PROGRESS
IN BIOMEDICAL
RESEARCH.”**

—Francis Collins

Director of the National
Institutes of Health (NIH), and

Kathy Hudson

NIH Deputy Director for
Science, Outreach, and Policy

It's especially important to gather as much patient data as possible in the new era of precision medicine, now that common diseases have been broken down into many diverse subtypes—each essentially its own rare disease.

But precision medicine not only requires massive datasets; it also demands free and easy access to that data. That access is being held back by a range of hurdles, stifling research and ultimately costing lives.

“Sharing data is essential for progress in biomedical research,” wrote Francis Collins, director of the National Institutes of Health (NIH), and Kathy Hudson, NIH deputy director for science, outreach, and policy, in a 2016 *New England Journal of Medicine* article. “Rapid data sharing was key to the success of the Human Genome Project, and that same commitment has been spreading across biomedicine in the past two decades, as advances in technology and ‘big data’ have enabled an entirely new level of data sharing and inquiry.”

Despite that progress, most of the genomic data being collected remains hoarded by the various institutions that compile it. Even academic centers, under a mandate to disseminate their findings, share only 50 percent of their data¹.

“There's a perception that there are no business incentives to share data,” said Bryce Olson, Intel's global marketing director for health and life sciences.

Scientific careers are made with peer-reviewed publications, and researchers compete with each other for space in prestigious journals. To many, sharing data could mean giving up the chance for a once-in-a-lifetime headline publication in *Science* or *Nature*.

The private institutions that employ many of these researchers also compete with each other for business—i.e., patients—and their discoveries become the competitive advantage that drives more patients to their centers.

¹ JAMA Oncology, November 2016

**“THERE'S A PERCEPTION
THAT THERE ARE NO
BUSINESS INCENTIVES
TO SHARE DATA.”**

—Bryce Olson
Global Marketing Director for
Health and Life Sciences, Intel

Regulators and industry experts say this is a shortsighted view that needs to change.

“Now that diseases have been hyper-segmented into so many unique types, hospitals are at serious risk of never having enough data on their own to get meaningful data about a given patient segment,” said Olson.

Olson knows this personally. His advanced prostate cancer was driven by a mutant, hyperactive molecular pathway that he successfully shut down for two years with a clinical trial drug targeting that pathway. He is likely one of a handful of people in the nation, possibly the world, to follow a similar path.

Only by sharing will any institution have enough data to move science forward and help more patients like him.

Challenges abound

Other obstacles contribute to data hoarding, including privacy concerns. Healthcare leaders always worry about privacy laws and typically err on the side of keeping data close, though shared genomic data typically removes identifiable patient information.

And then there's money. Even if everyone agreed to share, organizations would have to allocate budget to hire people and dedicate IT resources.

“Attend a panel at any precision medicine conference and you'll hear leaders talk about how they know they should be sharing,” said Olson. “But when they look at competing priorities at their own institutions, it's hard to convince people that this is important enough to spend that much money on it.”

Technical and cultural issues also impede data sharing. Researchers at the same institution may simply not know that their colleagues have relevant data. Unwieldy storage, formatting, and access procedures can make data difficult to find and retrieve even within the same organization. All these issues can take so much time that data may be outdated by the time a researcher finally gets it.

Sharing solutions

It will take a combination of mandates, industry changes, and patient activism to solve the problem. The federal government has taken some recent steps: The 21st Century Cures Act, passed in 2015, allows the NIH to require that data generated from NIH-supported research be shared.

“This gives all scientists the opportunity to use these data as quickly as possible to advance biomedical research,” said Olson.

Experts say other federal agencies need to do more, and that greater harmony is needed among the data-sharing regulations of all agencies, including the Food and Drug Administration, the Department of Health and Human Services, and the

**“IF YOU COMPARE
HEALTHCARE TO
OTHER INDUSTRIES,
IT'S BEEN A BIT
BEHIND FROM
THE PERSPECTIVE
OF DIGITAL
TRANSFORMATION.
SO IT'S VERY EXCITING
TO SEE THAT WE ARE
FINALLY HARNESSING
DATA IN NEW WAYS.”**

—Jennifer Esposito
Worldwide General Manager,
Health and Life Sciences, Intel

Health Insurance Portability and Accountability Act (HIPAA).

For now, providers or anyone looking to increase data sharing in a particular area will first need to understand the organizations holding the data to bring them together and answer the bigger questions. Those groups will have a vested interest in getting answers, said Olson. That will determine who to work with and who should chip in to cover the costs of a collaborative project.

“Sometimes you have to spend a lot of time just to educate data holders about what they have in terms of value,” said Olson, who has heard from many precision medicine project leaders in hospitals about how they've been successful. “Determine who is really motivated to solve a problem and explain to them that including their data will enable a better overall result, because the more you share, the more you actually have.”

Organizations shouldn't worry about losing a competitive edge by sharing, Olson added. “The intellectual property isn't the data. The data is the data. The IP comes from what each institution can find by analyzing the collective pool of data that is much greater than they would have working in a silo.”

Next is finding the funding to make it happen. Patient groups are a good place to start; they're motivated and organized, often have existing connections, and may have industry sponsors and partnerships to provide financial support.

“The money to make data sharing part of a project is often an afterthought,” said Olson. “Grant writers should build money for data sharing into their initial proposals, especially if they're applying for government-funded grants.”

Additional funding can come from patient advocacy groups with a vested interest in advancing research for a given condition. The Multiple Myeloma Foundation, for example, was founded by Kathy Guisti after she was diagnosed with multiple myeloma in 1998, when no effective treatments existed. The foundation created its own end-to-end ecosystem, including a tissue bank and genomics study, and put all the data in the public domain.

It's also important to pick cases that can really impact care, instead of broad data aggregation efforts without knowing the patients that will benefit. For example, it's often easier to motivate people to share data and raise funds for rare conditions or ones that lack good treatment options, as was the case with multiple myeloma in the late 1990s.

Build the right infrastructure for data

Once the data is found and funding secured, the technological challenges still need to be addressed. Datasets need to be accessible, standardized with a common vocabulary, and reusable (i.e., shared via open licensing).

But how does one grant access to the data? Moving large datasets is not feasible.

"This is where a federated model is so important," said Jennifer Esposito, Intel's worldwide general manager, health and life sciences. "You get the benefits of sharing without the problems of actually having to move data from one place to the other."

Healthcare providers will need to closely track the still-evolving data integration and classification standards. Until the industry settles on common standards, data center architectures must allow for flexibility and integration.

"That's why it's critical to build data center infrastructure that can help healthcare move to the cloud securely and efficiently," said Esposito. "As healthcare digitally transforms, we have to think about building the right foundation to scale out over the course of the future."

Consent policies also need to improve to facilitate more effective data sharing.

"The irony with our consent rules today is that patients don't want it hoarded," said Olson. "Patient genomic data is ultimately owned by the folks who provided the samples to get there—the patients—and they don't want it hidden away."

Getting those patients educated and involved is the real key to making progress.

"One way to drive more data sharing is making patients aware that their data is extremely valuable and holds the keys to unlock new insights," said Olson. "We need to help patients see the value when we ask their consent for greater use of their data," said Olson. "Patients want to know if a new trial opens up offering access to a drug that is a great molecular match for them. The effective, collaborative use of their data can make that happen."

And the technology underpinning it all needs to be easy to use and accessible.

"Ultimately, institutions need to figure out how to collect and access data that is much greater than they could ever hold individually, and enable this to happen easily, said Olson. "We really need to make this as easy as an online banking transaction for the physicians trying to bring the right treatments to the right patients."

Breaking down these business and technological hurdles will allow precision medicine to move into the fast lane.

"I've been in healthcare for about 20 years, but I've also been a patient," said Esposito. "And in both of those cases I've seen the technology really coming into play. If you compare healthcare to other industries, it's been a bit behind from the perspective of digital transformation. So it's very exciting to see that we are finally harnessing data in new ways. The next step is to learn how to share it."

